

A VARIATIONAL FORMULA FOR RISK-SENSITIVE REWARD

V. ANANTHARAM¹ and V. S. BORKAR²

ABSTRACT: We derive a variational formula for the optimal growth rate of reward in the infinite horizon risk-sensitive control problem for discrete time Markov decision processes with compact metric state and action spaces, extending a formula of Donsker and Varadhan for the Perron-Frobenius eigenvalue of a positive operator. This leads to a concave maximization formulation of the problem of determining this optimal growth rate.

Key words: risk-sensitive control; Perron-Frobenius eigenvalue; positive operators; variational formula

¹EECS Department, University of California, Berkeley, CA 94720, USA. Research supported in part by the ARO MURI grant W911NF-08-1-0233, Tools for the Analysis and Design of Complex Multi-Scale Networks, the NSF grants CNS-0910702 and ECCS-1343398, and the NSF Science & Technology Center grant CCF-0939370, Science of Information. A part of this work was done while this author was visiting IIT Bombay.

²Department of Elec. Engg., IIT Bombay, Powai, Mumbai 400076, India. Work supported in part by a J. C. Bose Fellowship and grant 11IRCCSG014 from IIT Bombay. A part of this work was done while this author was visiting the University of California, Berkeley.

1 Introduction

Infinite time horizon risk-sensitive control seeks to maximize the asymptotic growth rate for mean multiplicative reward in the standard Markov decision theory setting. The optimal reward multiplier per step turns out to be the Perron-Frobenius eigenvalue of a positive 1-homogeneous nonlinear operator. The existence of this Perron-Frobenius eigenvalue and an associated eigenfunction is ensured by the nonlinear Krein-Rutman theorem of [37, Theorem 3.1.1 and Proposition 3.1.5] under suitable conditions (see also [36], [33], [32], [12], [3]). Our aim here is to build on this nonlinear Krein-Rutman theorem to provide a variational formula for the optimal growth rate of reward in the spirit of the Donsker-Varadhan formula for the Perron-Frobenius eigenvalue of a nonnegative matrix [15, section 3.1.2], [18], [22].

Risk-sensitive control has traditionally been studied in the framework of cost minimization, see e.g. [16], [26], [27] for recent work on general state space models and [20], [24] for its discrete state space precursors. Work on risk-sensitive reward maximization has been relatively uncommon, see e.g. [28]. Unlike in the case of the classical discounted or ergodic costs, the two risk-sensitive control problems are not trivially equivalent by treating cost as a negative reward. In fact, risk-sensitive reward maximization is the natural set-up in portfolio optimization, see e.g. [13]. Nevertheless, it has been commonplace to replace it by risk-sensitive cost minimization so as to exploit the vastly more abundant available machinery for the latter problem, see, e.g. equation (18) of [6]. Interestingly, our approach is tailored for the risk-sensitive reward maximization problem.

The paper is organized as follows. This section presents the basic notation and control-theoretic framework. In section 2 we develop the role of the nonlinear Krein-Rutman theorem in giving an expression for the optimal reward multiplier per stage. In section 3 this is parlayed into a variational expression for the optimal growth rate of reward. Theorem 4 in section 3 is the main result of this paper. Alternative variational formulations derived from the primary one are discussed in section 4; each of these provides a different kind of insight into how to think about the optimal growth rate of reward. Some examples are worked out in section 5 to illustrate the nature of the results. We close the paper with some concluding remarks in section 6.

We turn next to introducing our notation and the control-theoretic framework. For a compact metric space \mathcal{X} , $\mathcal{M}(\mathcal{X})$ and $\mathcal{P}(\mathcal{X})$ will denote respectively the space of finite (signed) Borel measures on \mathcal{X} and the space of probability measures on \mathcal{X} , both with the topology of weak convergence [9]. $C(\mathcal{X})$ will denote the Banach space of continuous maps $\mathcal{X} \mapsto \mathcal{R}$ with the supremum norm, denoted by $\|\cdot\|$. Thus $\mathcal{M}(\mathcal{X})$ is the dual Banach space of $C(\mathcal{X})$, with the weak-* topology [39]. Let \mathcal{S} be a prescribed compact metric space called the *state space* and U another compact metric space, called the *action space*. We shall consider an \mathcal{S} -valued controlled Markov process $(X_n, n \geq 0)$ controlled by a U -valued control process $(Z_n, n \geq 0)$ defined as follows. Consider a complete probability space (Ω, \mathcal{F}, P) where $\Omega := (\mathcal{S} \times U)^\infty$, and \mathcal{F} is its product Borel σ -field. For $\omega = [(\omega_0, \omega'_0), (\omega_1, \omega'_1), (\omega_2, \omega'_2), \dots] \in \Omega$ with $\omega_i \in \mathcal{S}$ and $\omega'_i \in U \ \forall i$, define ‘canonical’ random variables $X_i = \omega_i, Z_i = \omega'_i, i \geq 0$. The probability measure P on (Ω, \mathcal{F}) is then the law of $((X_n, Z_n), n \geq 0)$ defined as follows. The law of X_0 is prescribed and the law of $((X_n, Z_n), n \geq 0)$ is constructed inductively. For this purpose, define two increasing families of sub- σ -fields of \mathcal{F} : $\mathcal{F}_n^- := \sigma(X_m, m \leq n; Z_m, m < n)$ and $\mathcal{F}_n := \sigma(X_m, m \leq n; Z_m, m \leq n)$ for $n \geq 0$. First define the conditional law of Z_0 given \mathcal{F}_0^- as $\phi_0(du|X_0)$, where

$$\phi_0(du|x_0) : \mathcal{S} \mapsto \mathcal{P}(U)$$

is a prescribed kernel, i.e. $\phi_0(du|x)$ is a probability distribution in $\mathcal{P}(U)$ for all x and $\phi_0(A|x)$ is Borel measurable in x for all Borel subsets $A \subset U$. Let P_n denote the law of $((X_0, Z_0), (X_1, Z_1), \dots, (X_n, Z_n))$, defined as a probability measure on (Ω, \mathcal{F}_n) , starting with $n = 0$. Define the law of X_{n+1} given \mathcal{F}_n as $p(dy|X_n, Z_n)$ where

$$p(dy|x, u) : \mathcal{S} \times U \mapsto \mathcal{P}(\mathcal{S})$$

is a prescribed kernel, i.e. $p(dy|x, u)$ is a probability distribution in $\mathcal{P}(\mathcal{S})$ for all $(x, u) \in \mathcal{S} \times U$ and $p(A|x, u)$ is Borel measurable in (x, u) for all Borel subsets $A \subset \mathcal{S}$. Define the conditional law of Z_{n+1} given \mathcal{F}_{n+1}^- as

$$\phi_{n+1}(du|(X_0, Z_0), \dots, (X_n, Z_n), X_{n+1})$$

where

$$\phi_{n+1}(du|(x_0, u_0), \dots, (x_n, u_n), x_{n+1}) : (\mathcal{S} \times U)^n \times \mathcal{S} \mapsto \mathcal{P}(U)$$

is a prescribed kernel for each n . These together define P_{n+1} . By the Ionescu-Tulcea theorem (p. 101, [38]), we define a unique P on (Ω, \mathcal{F}) . By construction, for all Borel $A \subset \mathcal{S}$,

$$\begin{aligned} P(X_{n+1} \in A | \mathcal{F}_n) &= P(X_{n+1} \in A | X_n, Z_n) \\ &= p(A | X_n, Z_n). \end{aligned} \tag{1}$$

The $(Z_n, n \geq 0)$ constructed above will be referred to as admissible controls. We shall also consider two special classes of admissible controls: *stationary Markov controls* of the form

$$Z_n = v(X_n) \quad \forall n,$$

for some measurable $v : \mathcal{S} \mapsto U$, and *randomized stationary Markov controls* satisfying

$$P(Z_n \in A | \mathcal{F}_n) = P(Z_n \in A | X_n) = \varphi(A | X_n) \quad \forall n, \forall \text{ Borel } A \subset U,$$

for some kernel $\varphi(du|x) : \mathcal{S} \mapsto \mathcal{P}(U)$. By a standard abuse of terminology, we identify these with the maps $v(\cdot), \varphi(\cdot|\cdot)$ resp. The sets thereof will be denoted by SM and RM respectively. We view SM as a subset of RM by identifying $v(\cdot)$ with $\delta_{v(\cdot)}$, the Dirac measure at $v(\cdot)$.

The infinite horizon risk-sensitive reward we seek to characterize is

$$\lambda := \sup_{x \in \mathcal{S}} \sup \liminf_{N \uparrow \infty} \frac{1}{N} \log E \left[e^{\sum_{m=0}^{N-1} r(X_m, Z_m, X_{m+1})} | X_0 = x \right], \tag{2}$$

where the second supremum is over all admissible controls. Here $r(x, u, y)$ is an extended-real-valued function on $\mathcal{S} \times U \times \mathcal{S}$, called the ‘per stage reward’ on transitioning from x to y under action u . It should be noted that we will allow $e^{r(x, u, y)} = 0$ for some (x, u, y) , so $r(x, u, y)$ should be thought of as being allowed to take the extended real value $-\infty$.

Throughout the paper, we make the following assumptions about $r(x, u, y)$ and $p(dy|x, u)$. We will occasionally explicitly recall these assumptions to remind the reader of this.

(A0): $e^{r(x, u, y)} \in C(\mathcal{S} \times U \times \mathcal{S})$.

(A1): The maps $(x, u) \mapsto \int f(y)p(dy|x, u)$, $f \in C(\mathcal{S})$ with $\|f\| \leq 1$, are equicontinuous. This is true, e.g., if \mathcal{S} is a compact metric space, U is a compact metric space, and $p(dy|x, u) = \psi(y|x, u)\varphi(dy)$ with $\varphi \in \mathcal{P}(\mathcal{S})$ having full support and $\psi(y|\cdot, \cdot), y \in \mathcal{S}$, equicontinuous.

We shall denote by e^{r_M} the least upper bound for $e^{r(\cdot, \cdot)}$, which is finite by virtue of assumption **(A0)**.

Towards the end of the next section we will build up to the main variational formula by first considering the case where we have additional restrictions captured by the following assumptions.

(A0+): Condition **(A0)** holds and we have $e^{r(x, u, y)} > 0$ for all (x, u, y) .

(A1+): Condition **(A1)** holds and $p(dy|x, u)$ has full support for all x, u . For instance, if \mathcal{S} is a compact metric space, U is a compact metric space, and $p(dy|x, u) = \psi(y|x, u)\varphi(dy)$ as above with $\psi(y|\cdot, \cdot), y \in \mathcal{S}$, equicontinuous, then $\psi(\cdot|x, u) > 0$ on \mathcal{S} will ensure that this assumption holds.

We shall denote by $e^{r_m} > 0$ the greatest lower bound for $e^{r(\cdot, \cdot)}$ when **(A0+)** holds.

If $p(dx)$ and $q(dx)$ are finite nonnegative Borel measures on a compact metric space \mathcal{X} , we write $D(p(dx)||q(dx))$ for the relative entropy of $p(dx)$ with respect to $q(dx)$, defined by

$$D(p(dx)||q(dx)) := \begin{cases} \int p(dx) \log l(x) & \text{if we can write } p(dx) = l(x)q(dx) \\ \infty & \text{otherwise.} \end{cases}$$

See e.g. [41] for some of the basic properties of relative entropy.

2 The Perron-Frobenius eigenvalue

Let assumptions **(A0)** and **(A1)** be in force. Define the operator $T : C(\mathcal{S}) \mapsto C(\mathcal{S})$ by

$$Tf(x) := \sup_{\phi \in \mathcal{P}(U)} \int \int p(dy|x, u)\phi(du)e^{r(x, u, y)}f(y). \quad (3)$$

For fixed $x \in \mathcal{S}$ on the left hand side of (3) the supremum on the right hand side is the expectation of a continuous affine function on a compact set of probability measures. Hence, it is a maximum attained at a Dirac measure. For each fixed $f \in C(\mathcal{S})$, a standard measurable selection theorem [5, Lemma 1, p. 182] allows us to choose the family of maximizers, parametrized by $x \in \mathcal{S}$, as a measurable function $v : \mathcal{S} \mapsto U$. To see that T is a map $C(\mathcal{S}) \mapsto C(\mathcal{S})$, note that for $f \in C(\mathcal{S})$ with $\|f\| \leq R$,

$$\begin{aligned}
& |Tf(x) - Tf(x')| \\
&= \left| \sup_{\phi \in \mathcal{P}(U)} \int \int p(dy|x, u) \phi(du) e^{r(x, u, y)} f(y) \right. \\
&\quad \left. - \sup_{\phi \in \mathcal{P}(U)} \int \int p(dy|x', u) \phi(du) e^{r(x', u, y)} f(y) \right| \\
&= \left| \sup_u \int p(dy|x, u) e^{r(x, u, y)} f(y) \right. \\
&\quad \left. - \sup_u \int p(dy|x', u) e^{r(x', u, y)} f(y) \right| \\
&\leq e^{r_M} \sup_u \sup_{f: \|f\| \leq R} \left| \int p(dy|x, u) f(y) \right. \\
&\quad \left. - \int p(dy|x', u) f(y) \right| + R \max_{u, y} \left| e^{r(x, u, y)} - e^{r(x', u, y)} \right|.
\end{aligned}$$

As $x \rightarrow x'$, the first term on the right tends to zero by **(A1)** and the second term on the right tends to zero by uniform continuity of e^r , being a continuous function defined on a compact set, by **(A0)**. In fact, this shows that $Tf, \|f\| \leq R$, are equicontinuous and bounded. Also, from the definition of T , it is straightforward to check that

$$\|Tf - Tg\| \leq e^{r_M} \|f - g\|.$$

which establishes T as a continuous (in fact, Lipschitz) map $C(\mathcal{S}) \mapsto C(\mathcal{S})$.

Likewise, define, for $f \in C(\mathcal{S})$,

$$T^{(n)}f(x) := \sup E \left[e^{\sum_{m=0}^{n-1} r(X_m, Z_m, X_{m+1})} f(X_n) | X_0 = x \right],$$

where the supremum is over all admissible control processes. Then $T^{(1)} = T$, by virtue of the measurable selection theorem alluded to after (3). We use

the convention $T^{(0)} :=$ the identity map.

Lemma 1. $(T^{(n)}, n \geq 0)$ is a semigroup of operators on $C(\mathcal{S})$. \square

Proof of Lemma 1: Note that we need to verify that $T^{(n)}$ for $n \geq 2$ maps $C(\mathcal{S})$ to $C(\mathcal{S})$ as part of the stated claim. This follows as a corollary of the proof, which establishes that $T^{(n)}$ is the n -fold concatenation of T with itself. The proof follows by a standard dynamic programming argument. Specifically, we first have

$$\begin{aligned}
& T^{(n)}f(x) \\
&= \sup E \left[e^{\sum_{m=0}^{n-1} r(X_m, Z_m, X_{m+1})} f(X_n) | X_0 = x \right] \\
&\leq \sup E \left[e^{r(X_0, Z_0, X_1)} \sup E \left[e^{\sum_{m=1}^{n-1} r(X_m, Z_m, X_{m+1})} f(X_n) | X_0, Z_0, X_1 \right] | X_0 = x \right] \\
&= \sup E \left[e^{r(X_0, Z_0, X_1)} T^{(n-1)}f(X_1) | X_0 = x \right], \tag{4}
\end{aligned}$$

where the inner supremum in the second line is over the control sequence from time 1 onwards, conditioned on $X_0 = x_0, Z_0 = z_0, X_1 = x_1$ (say). Secondly, let $\epsilon > 0$. By [10, Lemma 1, p. 55], conditioned on (X_0, Z_0, X_1) , there exists an admissible state-control sequence $(X'_m, Z'_m), m \geq 1$, with $X'_1 = X_1$ such that

$$\begin{aligned}
& E \left[e^{\sum_{m=1}^{n-1} r(X'_m, Z'_m, X'_{m+1})} f(X'_n) | X'_1 \right] \\
&\geq \sup E \left[e^{\sum_{m=1}^{n-1} r(X_m, Z_m, X_{m+1})} f(X_n) | X_1 \right] - \epsilon, \text{ a.s.}
\end{aligned}$$

Let $X'_0 = X_0 = x, Z'_0 := \operatorname{argmax}(\int p(dy|x, \cdot) e^{r(x, \cdot, y)} T^{(n-1)}f)$. Then

$$(X'_0, Z'_0), (X'_1, Z'_1), \dots, (X'_n, Z'_n))$$

is an admissible state-control sequence and

$$\begin{aligned}
T^{(n)}f(x) &\geq E \left[e^{\sum_{m=0}^{n-1} r(X'_m, Z'_m, X'_{m+1})} f(X'_n) | X'_0 = x \right] \\
&\geq E \left[e^{r(X_0, Z_0, X_1)} \sup E \left[e^{\sum_{m=1}^{n-1} r(X_m, Z_m, X_{m+1})} f(X_n) | X_1 \right] \right] - e^{r_M} \epsilon \\
&= E \left[e^{r(X_0, Z_0, X_1)} T^{(n-1)}f(X_1) | X_0 = x \right] - e^{r_M} \epsilon \\
&= T^{(1)}(T^{(n-1)}f)(x) - e^{r_M} \epsilon \tag{5}
\end{aligned}$$

Combining (4), (5) and using the fact that $\epsilon > 0$ was arbitrary, it follows that $T^{(n)} = T^{(1)} \circ T^{(n-1)}$. A similar argument shows that $T^{(n)}f = T^{(n-1)} \circ Tf$. \square

The semigroup $(T^{(n)}, n \geq 0)$, is precisely the discrete time counterpart of the Nisio semigroup [35].

Let $C^+(\mathcal{S}) := \{f \in C(\mathcal{S}) : f(x) \geq 0\}$ denote the set of nonnegative functions in $C(\mathcal{S})$. Then $C^+(\mathcal{S})$ is a *cone*, i.e. it is closed under addition and scalar multiplication by nonnegative real numbers, and we have $C^+(\mathcal{S}) \cap (-C^+(\mathcal{S})) = \{\theta\}$ where θ denotes the constant function that is identically zero. Thus $C^+(\mathcal{S})$ defines a partial order on $C(\mathcal{S})$, denoted \geq , given by $f \geq g$ if $f - g \in C^+(\mathcal{S})$. We write $f > g$ (equivalently, $g < f$) if $f \geq g, f \neq g$, and we write $f >> g$ if $f - g$ is a strictly positive function in $C(\mathcal{S})$ or equivalently if $f - g \in \text{int}(C^+(\mathcal{S}))$, where $\text{int}(C^+(\mathcal{S}))$ denotes the interior of $C^+(\mathcal{S})$. The *dual cone* of $C^+(\mathcal{S})$ is the cone in the dual Banach space $\mathcal{M}(\mathcal{S})$ given by $\{\mu \in \mathcal{M}(\mathcal{S}) : \int f d\mu \geq 0 \forall f \in C^+(\mathcal{S})\}$. This is the set of finite nonnegative measures on \mathcal{S} , which we denote by $\mathcal{M}^+(\mathcal{S})$. For more on cones in Banach spaces, see [2].

Let us now make the additional assumption **(A0+)** and **(A1+)**. One can then verify the following additional properties of $T^{(n)}$ for each $n \geq 1$.

1. $T^{(n)}$ is strictly increasing, i.e., $f < g$ implies $T^{(n)}f < T^{(n)}g$. In view of the fact established above that $(T^{(n)}, n \geq 0)$ is a semigroup, it suffices to prove this claim for $n = 1$. We know that there is a measurable function $v : \mathcal{S} \mapsto U$ such that

$$Tf(x) = \int p(dy|x, v(x))e^{r(x, v(x), y)} f(y) .$$

Then

$$\begin{aligned} & Tg(x) - Tf(x) \\ & \geq \int p(dy|x, v(x))e^{r(x, v(x), y)} g(y) - \int p(dy|x, v(x))e^{r(x, v(x), y)} f(y) \\ & \geq e^{nr_m} \int p(dy|x, v(x))(g(y) - f(y)) \\ & > 0 , \end{aligned}$$

because $f < g$, $f \neq g$ and $\text{support}(p(dy|x, u)) = \mathcal{S} \forall x, u$.

2. $T^{(n)}$ is strongly positive, i.e., $f \in C^+(\mathcal{S})$, $f \neq \theta \implies T^{(n)}f \in \text{int}(C^+(\mathcal{S}))$.
This follows from the fact that for any $u_0 \in U$,

$$T^{(n)}f(x) \geq e^{nr_m} \int p(dy|x, u_0)f(y) > 0,$$

where we use the fact that $\text{support}(p(dy|x, u_0)) = \mathcal{S}$.

3. $T^{(n)}$ is positively one-homogeneous, i.e., for $c > 0$, $T^{(n)}(cf) = cT^{(n)}f$.
(This holds under the weaker assumptions **(A0)** and **(A1)**.)
4. For $M > e^{-nr_m}$ and $\check{f} \in C(\mathcal{S})$ defined by $\check{f}(\cdot) \equiv 1$, $MT^{(n)}\check{f} > \check{f}$.
5. $T^{(n)}$ is compact. (This holds under the weaker assumptions **(A0)** and **(A1)**.) It suffices to verify this for $n = 1$, the general case being then a consequence of the semigroup property. By **(A1)**, the family $x \mapsto F_f(x, u) := \int f(y)e^{r(x, u, y)}p(dy|x, u)$, $u \in U$, $\|f\| \leq R$, is equicontinuous and bounded in $C(\mathcal{S})$ -norm by $e^{r_M}R$. Hence it is relatively compact in $C(\mathcal{S})$ by the Arzela-Ascoli theorem. Let $\delta \in [0, 1] \mapsto w_\delta(\cdot)$ denote its common modulus of continuity relative to a compatible metric κ on \mathcal{S} . Then $T : C(\mathcal{S}) \mapsto C(\mathcal{S})$ satisfies $\|Tf\| \leq e^{r_M}R$ for $\|f\| \leq R$, $f \in C(\mathcal{S})$, and,

$$\begin{aligned} & \sup_{x, y \in \mathcal{S}, \kappa(x, y) < \delta} \|Tf(x) - Tf(y)\| \\ & \leq \sup_{x, y \in \mathcal{S}, \kappa(x, y) < \delta} \left\| \sup_u F_f(x, u) - \sup_u F_f(y, u) \right\| \\ & \leq \sup_{x, y \in \mathcal{S}, \kappa(x, y) < \delta} \sup_u \|F_f(x, u) - F_f(y, u)\| \\ & \leq w_\delta(F_f) \xrightarrow{\delta \downarrow 0} 0 \end{aligned}$$

uniformly in $f : \|f\| \leq R$. Thus $Tf, \|f\| \leq R$, ie equicontinuous. By Arzela-Ascoli theorem, it is relatively compact, implying that $T : C(\mathcal{S}) \mapsto C(\mathcal{S})$ is a compact operator.

The preceding considerations allow us to state the following theorem.

Theorem 1. *Under the assumptions **(A0+)** and **(A1+)**, there exists a unique $\rho > 0$ (the Perron-Frobenius eigenvalue) and a $\psi \in \text{int}(C^+(\mathcal{S}))$ such that $T\psi = \rho\psi$, i.e.,*

$$\rho\psi(x) = \sup_{\phi \in \mathcal{P}(U)} \int \int p(dy|x, u) \phi(du) e^{r(x, u, y)} \psi(y), \quad (6)$$

with ρ given by

$$\begin{aligned} \rho &= \inf_{f \in \text{int}(C^+(\mathcal{S}))} \sup_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu} \\ &= \sup_{f \in \text{int}(C^+(\mathcal{S}))} \inf_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu}. \end{aligned} \quad (7)$$

□

Equation (7) is an abstract version of the celebrated Collatz-Wielandt formula for the Perron-Frobenius eigenvalue of irreducible nonnegative matrices, see e.g. [34].

Before proceeding to the proof of Theorem 1, it is appropriate to make a few remarks. A great deal is known about analogs of the Perron-Frobenius theorem for increasing positively one-homogeneous maps on finite dimensional vector spaces, see the recent book [30]. When the map is on an ordered Banach space (and one talks about a Krein-Rutman theorem rather than a Perron-Frobenius theorem, in view of the seminal work in [29]), we rely on Theorem 3.1.1, Proposition 3.1.5, and Lemma 3.1.7 of [37], as seen in the proof below (see also [36], [33]). These results in [37] are themselves stated in a much broader context than the special case of the Banach space $C(\mathcal{S})$ and the order structure defined by the cone $C^+(\mathcal{S})$, with \mathcal{S} a compact metric space, which suffices for our purposes. The recent papers [32] and [12] claim even stronger nonlinear Krein-Rutman theorems. However, it has been recognized in [3] that some of the claims in these papers are wrong. The proof of the Theorem 1 given below does not rely in any way on [32], [12], or [3].

Proof of Theorem 1: We define

$$\|T^{(n)}\|_+ := \sup\{\|T^{(n)}f\| : f \in C^+(\mathcal{S}), \|f\| \leq 1\}, \quad n \geq 0.$$

Since $(T^{(n)}, n \geq 0)$ is a positive semigroup, it is straightforward to check that $\|T^{(k+l)}\|_+ \leq \|T^{(k)}\|_+ \|T^{(l)}\|_+$ for all $k, l \geq 0$, and so

$$r(T) := \lim_{n \rightarrow \infty} \|T^{(n)}\|_+^{\frac{1}{n}}$$

exists. By the fourth of the properties of the semigroup $(T^{(n)}, n \geq 0)$ shown above, we have $r(T) > 0$. It will turn out that the ρ promised in the statement of Theorem 1 is just $r(T)$.

Strong positivity of T , which was shown above, verifies assumption A4 in [37, pg. 47], and the facts that T is compact (as established above), one-homogeneous, and order preserving are respectively the conditions A1, A2, and A3 in [37, pg.47]. Thus [37, Proposition 3.1.5.] provides the additional requirement in the statement of [37, Theorem 3.1.1] that T have an eigenvalue, and [37, Theorem 3.1.1] states that with ρ taken to be $r(T)$ there exists a $\psi \in \text{int}(C^+(\mathcal{S}))$ such that (6) holds.

It remains to establish (7), where we now know that $\rho = r(T)$. We have

$$\rho \geq \inf_{f \in \text{int}(C^+(\mathcal{S}))} \sup_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu},$$

which comes from substituting ψ as a choice for f on the right hand side. Similarly, we have

$$\rho \leq \sup_{f \in \text{int}(C^+(\mathcal{S}))} \inf_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu}.$$

Thus it suffices to establish

$$\inf_{f \in \text{int}(C^+(\mathcal{S}))} \sup_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu} \geq \rho \geq \sup_{f \in \text{int}(C^+(\mathcal{S}))} \inf_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu}. \quad (8)$$

Given $f \in \text{int}(C^+(\mathcal{S}))$, we have

$$Tf \leq \left(\sup_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu} \right) f.$$

From [37, Lemma 3.1.7 (ii)], we have $r(T) \leq \sup_{\mu \in \mathcal{M}^+(\mathcal{S})} \frac{\int T f d\mu}{\int f d\mu}$. Since this holds for all $f \in \text{int}(C^+(\mathcal{S}))$, this establishes the first inequality in (8). The

proof of the second inequality in (8) is similar, based on [37, Lemma 3.1.7 (iii)]. This concludes the proof of Theorem 1. \square

Next we show that $\log \rho$ is in fact the optimal growth rate of the risk-sensitive reward. For a development of the analogous result in the case of controlled diffusion processes, see [4]. As argued earlier, in connection with the right hand side of (3), for each $x \in \mathcal{S}$, the supremum on the right hand side of (6) is the expectation of a continuous affine function on a compact set of probability measures, and is therefore a maximum attained at a Dirac measure. A standard measurable selection theorem [5, Lemma 1, p. 182] then allows us to identify the family of maximizers, parametrized by $x \in \mathcal{S}$, with an element of SM , which we denote by $v^*(\cdot)$. Letting $(X_n^*, n \geq 0)$ denote the chain governed by the stationary Markov strategy $v^*(\cdot)$ and $(Z_n^* = v^*(X_n^*), n \geq 0)$ the corresponding control sequence, we then have

$$\rho\psi(x) = E \left[e^{r(x, v^*(x), X_1^*)} \psi(X_1^*) \right],$$

and, more generally, by iterating, we have, for all $x \in \mathcal{S}$,

$$\rho^n \psi(x) = E \left[e^{\sum_{m=0}^{n-1} r(X_m^*, Z_m^*, X_{m+1}^*)} \psi(X_n^*) | X_0^* = x \right].$$

Since $\psi(x) \in \text{int}(C^+(\mathcal{S}))$, we have $0 < c < \psi(\cdot) < C < \infty$ for some constants c, C when ψ is chosen with, say, $\|\psi\| = 1$. Thus, for all $x \in \mathcal{S}$,

$$\frac{c}{C} E \left[e^{\sum_{m=0}^{n-1} r(X_m^*, Z_m^*, X_{m+1}^*)} | X_0^* = x \right] \leq \rho^n \leq \frac{C}{c} E \left[e^{\sum_{m=0}^{n-1} r(X_m^*, Z_m^*, X_{m+1}^*)} | X_0^* = x \right].$$

Hence

$$\log \rho = \lim_{n \uparrow \infty} \frac{1}{n} \log E \left[e^{\sum_{m=0}^{n-1} r(X_m^*, Z_m^*, X_{m+1}^*)} | X_0^* = x \right].$$

For any other admissible state-control sequence $((X_n, Z_n), n \geq 0)$, we have

$$\rho\psi(x) \leq E \left[e^{r(x, Z_0, X_1)} \psi(X_1) | X_0 = x \right].$$

Iterating,

$$\rho^n \psi(x) \leq E \left[e^{\sum_{m=0}^{n-1} r(X_m, Z_m, X_{m+1})} \psi(X_n) | X_0 = x \right].$$

and therefore

$$\log \rho \leq \liminf_{n \uparrow \infty} \frac{1}{n} \log E \left[e^{\sum_{m=0}^{n-1} r(X_m, Z_m, X_{m+1})} | X_0 = x \right].$$

We have proved:

Theorem 2. *Under the assumptions **(A0+)** and **(A1+)**, we have, for all $x \in \mathcal{S}$,*

$$\log \rho = \sup \liminf_{n \uparrow \infty} \frac{1}{n} \log E \left[e^{\sum_{m=0}^{n-1} r(X_m, Z_m, X_{m+1})} | X_0 = x \right],$$

where the supremum on the right is over all admissible controls and ρ on the left is given as in Theorem 1. Furthermore, this supremum is a maximum attained at some $v^*(\cdot) \in SM$. \square

An immediate consequence is the following.

Corollary 1. *Under the assumptions **(A0+)** and **(A1+)** we have*

$$\lambda = \log \rho ,$$

where λ is the optimal growth rate of reward, as defined in (2), and ρ is as defined in Theorem 1. \square

3 A variational formula

By (7), we have

$$\begin{aligned} \rho &= \inf_{f \gg 0} \sup_{\mu \in \mathcal{M}^+(S): \int f d\mu = 1} \int \mu(dx) \sup_u \int p(dy|x, u) e^{r(x, u, y)} f(y) \\ &= \inf_{f \gg 0} \sup_{\nu \in \mathcal{P}(S)} \int \nu(dx) \left(\frac{\sup_u \int p(dy|x, u) e^{r(x, u, y)} f(y)}{f(x)} \right) \\ &= \inf_{f \gg 0} \sup_x \left(\frac{\sup_u \int p(dy|x, u) e^{r(x, u, y)} f(y)}{f(x)} \right) \\ &= \inf_{f \gg 0} \sup_x \sup_u \int p(dy|x, u) e^{r(x, u, y) + \log f(y) - \log f(x)} \\ &= \inf_{f \gg 0} \sup_{\gamma \in \mathcal{P}(\mathcal{S} \times U)} \int \int \gamma(dx, du) \int p(dy|x, u) e^{r(x, u, y) + \log f(y) - \log f(x)}. \end{aligned}$$

Introduce the notation

$$\begin{aligned}\eta(dx, du, dy) &= \eta_0(dx)\eta_1(du|x)\eta_2(dy|x, u) \\ &= \tilde{\eta}(dx, du)\eta_2(dy|x, u).\end{aligned}$$

Let

$$\mathcal{G} := \{\eta(dx, du, dy) : \eta_0 \text{ is invariant under the transition kernel } \int_U \eta_2(dy|x, u)\eta_1(du|x)\},$$

i.e. $\eta \in \mathcal{G}$ iff

$$\int \tilde{\eta}(dx, du)\eta_2(dy|x, u) = \eta_0(dy) .$$

Recall that $D(\cdot||\cdot)$ is convex and lower semi-continuous in both arguments [41]. Then

$\log \rho$

$$\begin{aligned}
&= \inf_{f > 0} \sup_{\gamma} \log \int \int \int \gamma(dx, du) p(dy|x, u) e^{r(x, u, y) + \log f(y) - \log f(x)} \\
&= \inf_{g \in C(\mathcal{S})} \sup_{\gamma} \log \int \int \int \gamma(dx, du) p(dy|x, u) e^{r(x, u, y) + g(y) - g(x)} \\
&= \inf_{g \in C(\mathcal{S})} \sup_{\gamma} \sup_{\eta} \int \int \int \eta(dx, du, dy) \left(r(x, u, y) + g(y) - g(x) \right) \\
&\quad - D(\eta(dx, du, dy) || \gamma(dx, du) p(dy|x, u)) \\
&\quad \text{(by the Gibbs variational formula (Prop. 1.4.2(a), pp. 33-34, [17])} \\
&= \sup_{\gamma} \sup_{\eta} \inf_{g \in C(\mathcal{S})} \int \int \int \eta(dx, du, dy) \left(r(x, u, y) + g(y) - g(x) \right) \\
&\quad - D(\eta(dx, du, dy) || \gamma(dx, du) p(dy|x, u)) \\
&\quad \dots\dots \text{(by the min-max theorem [19])} \\
&= \sup_{\gamma} \sup_{\eta} \inf_{g \in C(\mathcal{S})} \left(\int \int \int \eta(dx, du, dy) \left(r(x, u, y) + g(y) - g(x) \right) \right. \\
&\quad \left. - \left(D(\tilde{\eta}(dx, du) || \gamma(dx, du)) + \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right) \right) \\
&= \sup_{\eta} \inf_{g \in C(\mathcal{S})} \left(\int \int \int \eta(dx, du, dy) \left(r(x, u, y) + g(y) - g(x) \right) \right. \\
&\quad \left. - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right. \\
&\quad \left. \dots\dots \text{(by setting } \gamma = \tilde{\eta}) \right) \\
&= \sup_{\eta \in \mathcal{G}} \left[\inf_{g \in C(\mathcal{S})} \left(\int \int \int \eta(dx, du, dy) \left(r(x, u, y) + g(y) - g(x) \right) \right. \right. \\
&\quad \left. \left. - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right) \right] \\
&\quad \dots\dots \text{(because } [\dots] = -\infty \forall \eta \notin \mathcal{G}) \\
&= \sup_{\eta \in \mathcal{G}} \left(\int \int \int \eta(dx, du, dy) r(x, u, y) \right. \\
&\quad \left. - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right) \\
&\quad \dots\dots \text{(because } \eta \in \mathcal{G} \implies \int \eta(dx, du, dy) (g(y) - g(x)) = 0)
\end{aligned}$$

Thus we have:

Theorem 3. *Under the assumptions **(A0+)** and **(A1+)**, the optimal growth rate of reward λ , as defined in (2), has the variational characterization*

$$\lambda = \log \rho = \sup_{\eta \in \mathcal{G}} \left(\int \int \int \eta(dx, du, dy) r(x, u, y) - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right), \quad (9)$$

where ρ is defined as in Theorem 1. \square

The following result, which uses a limiting argument to strengthen Theorem 3, is the main result of this paper.

Theorem 4. *Under the assumptions **(A0)** and **(A1)**, the optimal growth rate of reward λ , as defined in (2), has the variational characterization*

$$\lambda = \sup_{\eta \in \mathcal{G}} \left(\int \int \int \eta(dx, du, dy) r(x, u, y) - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right). \quad (10)$$

\square

Before proving Theorem 4, let us first consider the uncontrolled case. We can fit this into our framework by taking U to be a set with one point, so that $p(dy|x, u) = \tilde{p}(dy|x)$ for all $u \in U$, for some kernel $\tilde{p}(dy|x)$, and $r(x, u, y) = \tilde{r}(x, y)$ for all $u \in U$, for some $\tilde{r}(\cdot, \cdot)$. Theorem 4 then specializes to the statement that the growth rate of the reward, under the respective specializations of conditions **(A0)** and **(A1)**, is given by

$$\lambda = \sup_{\alpha \in \tilde{\mathcal{G}}} \left(\int \int \int \alpha(dx, dy) \tilde{r}(x, y) - \int \int \alpha_0(dx) D(\alpha_1(dy|x) || \tilde{p}(dy|x)) \right)$$

where $\alpha(dx, dy) = \alpha_0(dx) \alpha_1(dy|x)$ and

$$\tilde{\mathcal{G}} := \{ \alpha(dx, dy) = \alpha_0(dx) \alpha_1(dy|x) : \int \alpha_0(dx) \alpha_1(dy|x) = \alpha_0(dy) \}.$$

This is then a version of the Donsker-Varadhan formula for the Perron-Frobenius eigenvalue of a positive operator [15], [18], [22].

Proof of Theorem 4: Let $\gamma(dy)$ be an arbitrary probability distribution on \mathcal{S} with full support, and, for all $\epsilon > 0$ sufficiently small, define the kernel

$$p_\epsilon(dy|x, u) := \frac{1}{a(x, u) + \epsilon} \left(e^{r(x, u, y)} p(dy|x, u) + \epsilon \gamma(dy) \right) ,$$

and the reward

$$r_\epsilon(x, u, y) := \log(a(x, u) + \epsilon) ,$$

where

$$a(x, u) := \int e^{r(x, u, y)} p(dy|x, u) .$$

Since this kernel and reward satisfy the conditions **(A0+)** and **(A1+)**, we have from Theorem 3 that the optimal growth rate of reward for the risk-sensitive reward maximization problem for this kernel and reward, call it λ_ϵ , is given by

$$\begin{aligned} \lambda_\epsilon = & \sup_{\eta \in \mathcal{G}} \left(\int \int \int \eta(dx, du, dy) r_\epsilon(x, u, y) \right. \\ & \left. - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p_\epsilon(dy|x, u)) \right) . \end{aligned} \quad (11)$$

From the formulation of the risk-sensitive objective we see that λ_ϵ is nondecreasing in ϵ , and that $\lambda_\epsilon \geq \lambda$ for all $\epsilon > 0$, where λ is defined as in (2). This can be seen by writing the expression for the n -step multiplicative reward, i.e.

$$E_\epsilon \left[e^{\sum_{m=0}^{N-1} r_\epsilon(X_m, Z_m, X_{m+1})} | X_0 = x \right] ,$$

as a multiple integral, which reveals that this quantity is monotonically non-decreasing in ϵ for any initial condition $x \in \mathcal{S}$ and any admissible control strategy. Thus $\lim_{\epsilon \rightarrow 0} \lambda_\epsilon$ exists and satisfies

$$\lim_{\epsilon \rightarrow 0} \lambda_\epsilon \geq \lambda . \quad (12)$$

To prove (10), we will first prove that

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \lambda_\epsilon \leq & \sup_{\eta \in \mathcal{G}} \left(\int \int \int \eta(dx, du, dy) r(x, u, y) \right. \\ & \left. - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right) , \end{aligned} \quad (13)$$

and then prove that

$$\begin{aligned} \lambda \geq & \sup_{\eta \in \mathcal{G}} \left(\int \int \int \eta(dx, du, dy) r(x, u, y) \right. \\ & \left. - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right). \end{aligned} \quad (14)$$

Together with (12), these two claims establish (10).

For fixed $\eta \in \mathcal{G}$, let $\Psi_\epsilon(\eta)$ denote the expression inside the outer brackets on the right hand side of (11). Then one has

$$\Psi_\epsilon(\eta) = - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || e^{r(x, u, y)} p(dy|x, u) + \epsilon \gamma(dy)) . \quad (15)$$

Similarly, for fixed $\eta \in \mathcal{G}$, let $\Psi_0(\eta)$ denote the expression inside the outer brackets on the right hand side of (10). We have

$$\Psi_0(\eta) = - \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || e^{r(x, u, y)} p(dy|x, u)) . \quad (16)$$

In fact, (15) reveals that for each $\eta \in \mathcal{G}$ we have $\Psi_\epsilon(\eta)$ is nondecreasing in ϵ , and together with (16), reveals that for all $\epsilon > 0$ and $\eta \in \mathcal{G}$, we have $\Psi_\epsilon(\eta) \geq \Psi_0(\eta)$. Thus we may conclude that for each $\eta \in \mathcal{G}$ the limit $\lim_{\epsilon \rightarrow 0} \Psi_\epsilon(\eta)$ exists, and that this limit satisfies $\lim_{\epsilon \rightarrow 0} \Psi_\epsilon(\eta) \geq \Psi_0(\eta)$.

Now, for all $\epsilon > 0$ and $\delta > 0$ sufficiently small, choose $\eta_\epsilon^\delta \in \mathcal{G}$ such that $\Psi_\epsilon(\eta_\epsilon^\delta) > \lambda_\epsilon - \delta$. Since \mathcal{G} is compact, there is a decreasing sequence $(\epsilon_m, m \geq 1)$ with $\lim_{m \rightarrow \infty} \epsilon_m = 0$, such that the sequence $(\eta_{\epsilon_m}^\delta, m \geq 1)$ has a limit in $\mathcal{P}(\mathcal{S} \times U \times \mathcal{S})$, call it η^δ . Further, since \mathcal{G} is closed, we have $\eta^\delta \in \mathcal{G}$. By the lower semicontinuity of $D(\cdot || \cdot)$ as a function of (\cdot, \cdot) [41] we have

$$\sup_{\eta \in \mathcal{G}} \Psi_0(\eta) \geq \Psi_0(\eta^\delta) \geq \lim_{m \rightarrow \infty} \Psi_{\epsilon_m}(\eta_{\epsilon_m}^\delta) \geq \lim_{m \rightarrow \infty} \lambda_{\epsilon_m} - \delta = \lim_{\epsilon \rightarrow 0} \lambda_\epsilon - \delta .$$

Since $\delta > 0$ was arbitrary, this establishes (13).

It remains to prove (14). If $\sup_{\eta \in \mathcal{G}} \Psi_0(\eta)$ (i.e. the right hand side of (14)) equals $-\infty$ then there is nothing to prove, so we may assume that this is not the case. Given $\eta \in \mathcal{G}$ for which $\Psi_0(\eta) \neq -\infty$, consider implementing the stationary Markov strategy defined by the kernel $\eta_1(du|x)$. The expected multiplicative reward after n steps when implementing this strategy, conditioned on starting with the initial distribution $\eta_0(dx_0)$, is

$$\int \cdots \int \eta_0(dx_0) \prod_{m=0}^{n-1} \eta_1(du_m|x_m) p(dx_{m+1}|x_m, u_m) e^{r(x_m, u_m, x_{m+1})} .$$

Since $\eta_2(dy|x, u)$ is absolutely continuous with respect to $p(dy|x, u)$ for almost all (x, u) , this equals

$$\int \cdots \int \eta_0(dx_0) \prod_{m=0}^{n-1} \eta_1(du_m|x_m) \eta_2(dx_{m+1}|x_m, u_m) e^{r(x_m, u_m, x_{m+1})} e^{-\log \frac{\eta_2(dx_{m+1}|x_m, u_m)}{p(dx_{m+1}|x_m, u_m)}}.$$

Let $\{X'_n\}$ denote a controlled Markov chain with controlled transition kernel $\eta_2(dy|x, u)$, initial law η_0 , and controlled by $\eta_1(du|x) \in RM$. Then

$$\begin{aligned} \lambda &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\int \cdots \int \eta_0(dx_0) \prod_{m=0}^{n-1} \eta_1(du_m|x_m) p(dx_{m+1}|x_m, u_m) e^{r(x_m, u_m, x_{m+1})} \right) \\ &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\int \cdots \int \eta_0(dx_0) \prod_{m=0}^{n-1} \eta_1(du_m|x_m) \eta_2(dx_{m+1}|x_m, u_m) \right. \\ &\quad \times \left. e^{r(x_m, u_m, x_{m+1}) - \log \frac{\eta_2(dx_{m+1}|x_m, u_m)}{p(dx_{m+1}|x_m, u_m)}} \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(E \left[e^{\sum_{m=0}^{n-1} (r(X'_m, Z'_m, X'_{m+1}) - \log \frac{d\eta_2(\cdot|X'_m, Z'_m)}{dp(\cdot|X'_m, Z'_m)}(X'_m))} \right] \right) \\ &\geq \lim_{n \rightarrow \infty} \frac{1}{n} E \left[\sum_{m=0}^{n-1} (r(X'_m, Z'_m, X'_{m+1}) - \log \frac{d\eta_2(\cdot|X'_m, Z'_m)}{dp(\cdot|X'_m, Z'_m)}(X'_m)) \right] \\ &\quad \text{(by Jensen's inequality)} \\ &= \Psi_0(\eta) \\ &\quad \text{(because } \eta \in \mathcal{G} \text{).} \end{aligned}$$

It follows that λ , as defined in (2), satisfies (14), which concludes the proof of Theorem 4. \square .

4 Remarks

1. Assume **(A0)**, **(A1)**. Fix $\varphi \in RM$, and consider $\{(X_n, Z_n), n \geq 0\}$ governed by the randomized stationary Markov strategy φ as an uncontrolled $\mathcal{S} \times U$ -valued Markov chain. To be precise, let $\bar{\mathcal{S}}$ denote $\mathcal{S} \times U$, let $\bar{U} := \{\bar{u}\}$ be a one point set, and define $\bar{p}(d\bar{y}|\bar{x}, \bar{u}) : \bar{\mathcal{S}} \times \bar{U} \mapsto \mathcal{P}(\bar{\mathcal{S}})$ by

$$\bar{p}(d\bar{y}|\bar{x}, \bar{u}) := p(dy|x, u) \varphi(du'|y),$$

where $\bar{x} := (x, u)$ and $\bar{y} := (y, u')$. Also, let

$$\bar{r}(\bar{x}, \bar{u}, \bar{y}) := r(x, u, y) .$$

It is straightforward to check that the assumptions **(A0)**, **(A1)** hold for the $\bar{\mathcal{S}}$ -valued chain with trivial control space \bar{U} and with the transition kernel and one step reward as above.

Given $\tau(dx, du, dy, du') = \tau_0(dx)\tau_1(du|x)\tau_2(dy|x, u)\tau_3(du'|x, u, y)$, write $\tilde{\tau}(dx, du)$ for $\tau_0(dx)\tau_1(du|x)$ and $\hat{\tau}(dy, du'|x, u)$ for $\tau_2(dy|x, u)\tau_3(du'|x, u, y)$. Let

$$\mathcal{G}_+ := \{ \tau(dx, du, dy, du') : \int \int \tilde{\tau}(dx, du) \hat{\tau}(dy, du'|x, u) = \tilde{\tau}(dy, du') \} .$$

Further, given $\tau(dx, du, dy, du')$, we define $\tau'(dx, du, dy, du')$ by setting

$$\tau'_0 := \tau_0, \quad \tau'_1 := \tau_1, \quad \tau'_2 := \tau_2, \quad \tau'_3(du'|x, u, y) := \tau_1(du'|y) ,$$

with the corresponding definitions for $\tilde{\tau}', \hat{\tau}'$. We claim that $\tau' \in \mathcal{G}_+$. To see this, first observe that $\int \int \tilde{\tau}(dx, du) \hat{\tau}(dy, du'|x, u) = \tilde{\tau}(dy, du')$ when integrated over u' gives $\int \int \tilde{\tau}(dx, du) \tau_2(dy|x, u) = \tau_0(dy)$. This means

$$\begin{aligned} \int \int \tilde{\tau}'(dx, du) \hat{\tau}'(dy, du'|x, u) &= \int \int \tilde{\tau}(dx, du) \tau_2(dy|x, u) \tau_1(du'|y) \\ &= \tau_0(dy) \tau_1(du'|y) \\ &= \tilde{\tau}(dy, du') = \tilde{\tau}'(dy, du') , \end{aligned}$$

which establishes the claim.

Let λ_φ denote the asymptotic growth rate of reward under the fixed randomized stationary Markov strategy φ . Then by applying Theorem 4 to the $\bar{\mathcal{S}}$ -valued chain with trivial control space \bar{U} defined above, we have

$$\begin{aligned} \lambda_\varphi &= \sup_{\tau \in \mathcal{G}_+} \left(\int \int \int \tau(dx, du, dy, U) r(x, u, y) - \right. \\ &\quad \left. \int \int \tilde{\tau}(dx, du) D(\hat{\tau}(dy, du'|x, u) || p(dy|x, u) \varphi(du'|y)) \right). \end{aligned} \quad (17)$$

Then we have

$$\begin{aligned}
& \sup_{\varphi} \lambda_{\varphi} \\
&= \sup_{\varphi \in RM} \sup_{\tau \in \mathcal{G}_+} \left(\int \int \int \tau(dx, du, dy, U) r(x, u, y) - \right. \\
&\quad \left. \int \int \tilde{\tau}(dx, du) D(\tau_2(dy|x, u) \tau_3(du'|x, u, y) || p(dy|x, u) \varphi(du'|y)) \right) \\
&\stackrel{(a)}{=} \sup_{\tau \in \mathcal{G}_+} \left(\int \int \int \tau'(dx, du, dy, U) r(x, u, y) - \right. \\
&\quad \left. \int \int \tilde{\tau}'(dx, du) D(\tau_2'(dy|x, u) || p(dy|x, u)) \right) \\
&\stackrel{(b)}{=} \sup_{\eta \in \mathcal{G}} \left(\int \int \int \eta(dx, du, dy) r(x, u, y) - \right. \\
&\quad \left. \int \int \tilde{\eta}(dx, du) D(\eta_2(dy|x, u) || p(dy|x, u)) \right) \\
&= \lambda.
\end{aligned}$$

Here, to justify step (a), notice that for every $\tau \in \mathcal{G}_+$, we have shown that $\tau' \in \mathcal{G}_+$. Therefore we have both

$$\int \int \int \tau'(dx, du, dy, U) r(x, u, y) = \int \int \int \tau(dx, du, dy, U) r(x, u, y)$$

and

$$\begin{aligned}
& \int \int \tilde{\tau}'(dx, du) D(\tau_2'(dy|x, u) || p(dy|x, u)) \\
&= \int \int \tilde{\tau}(dx, du) D(\tau_2(dy|x, u) || p(dy|x, u)).
\end{aligned}$$

The choice of $\varphi(du'|y) = \tau_1(du'|y)$ (which also equals $\tau_3'(du'|x, u, y)$) would make the expression

$$\int \int \int \tilde{\tau}'(dx, du) \tau_2'(dy|x, u) D(\tau_3'(du'|x, u, y) || \varphi(du'|y))$$

equal to zero, whereas the expression

$$\int \int \int \tilde{\tau}(dx, du) \tau_2(dy|x, u) D(\tau_3(du'|x, u, y) || \varphi(du'|y))$$

is nonnegative. To justify step (b) note that for every $\tau \in \mathcal{G}_+$, we have $\tau_0(dx)\tau_1(du|x)\tau_2(dy|x, u) \in \mathcal{G}$, and conversely for every $\eta \in \mathcal{G}$ we get $\tau \in \mathcal{G}_+$ by defining $\tau(dx, du, dy, du') := \eta(dx, du, dy)\eta_1(du'|y)$. Furthermore, this τ satisfies $\tau' = \tau$.

The upshot is that we have proved

$$\lambda = \sup_{\varphi \in RM} \lambda_\varphi. \quad (18)$$

Under **(A0+)**, **(A1+)**, this supremum is in fact a maximum by virtue of Theorem 2.

2. Since $D(\cdot|\cdot)$ is convex and lower semi-continuous in its arguments as noted earlier, (10) is a concave maximization problem on the convex³ set

$$\mathcal{G}_1 := \left\{ \eta(dx)\varphi(du|x)\mu(dy|x, u) : \eta \text{ is invariant under the transition kernel } x \mapsto \int_U \varphi(du|x)\mu(dy|x, u) \right\}.$$

It is worthwhile to compare this formulation with the classical dynamic programming approach. Recall that the dynamic programming equation (6) is the nonlinear eigenvalue problem

$$\rho V(x) = \sup_{\varphi} \left(\int \int p(dy|x, u)\varphi(dy|u)e^{r(x, u, y)}V(y) \right).$$

Consider the standard ‘log transformation’ $\zeta(x) := \log V$. Then

$$\log \rho + \zeta(x) = \sup_{\varphi} \log \left(\int \int p(dy|x, u)\varphi(du|x)e^{r(x, u, y) + \zeta(y)} \right).$$

We treat x as a fixed parameter on the right hand side. By the Gibbs variational principle, we have

$$\begin{aligned} & \log \rho + \zeta(x) \\ &= \sup_{\varphi} \sup_{\mu(\cdot, \cdot|x) \in \mathcal{P}(U \times \mathcal{S})} \left(\int \mu(du, dy|x)(r(x, u, y) + \zeta(y)) - \right. \\ & \quad \left. D(\mu(du, dy|x) || p(dy|x, u)\varphi(du|x)) \right). \end{aligned} \quad (19)$$

³See [11, section 11.2.3, p. 358] for the proof of convexity

Equation (19) is the dynamic programming equation for an ergodic team problem whose ‘per stage payoff’ function is

$$r(x, u, y) - D(\mu(du, dy|x) || p(dy|x, u)\varphi(du|x)),$$

where μ specifies an additional control variable the choice of which is in fact the distribution of the next state and control, whereas the original randomized control φ affects only the payoff. This is a team problem as opposed to a control problem because while both controls have the same objective, viz., to maximize a common reward, they are implemented in a non-cooperative manner. This is reminiscent of, e.g., [24], which considers the cost minimization formulation in which a similar procedure leads to a zero sum ergodic game. There does not, however, appear to be any corresponding development earlier for the reward maximization problem with a positive reward. While this is completely analogous to the game situation, we have obtained it without an explicit minorization condition as in [16], or the ‘condition B’ of [26]. We have instead conditions **(A0)** and **(A1)** which are relatively mild, and compactness of state space, which is not. We are working towards relaxing the latter.

An important point to note here is that we have an equivalent problem of maximizing a concave upper semi-continuous function over the convex set \mathcal{G}_1 . This is in contrast with the ergodic *team* problem of maximizing the same function over the *nonconvex* set

$$\mathcal{G}_2 := \{\eta(dx)\varphi(du|x)\mu(dy|x) : \eta \text{ is invariant under the transition kernel } x \mapsto \mu'(dy|x)\},$$

i.e., where the controls φ, μ' are chosen by the two team members non-cooperatively. The latter is what one obtains from the team formulation via log transformation.

3. It is also worth noting that the entropic penalty implicit in our variational formula also arises in different contexts [8], [23], [40].

5 Examples

5.1 Path counting on graphs

Let G be a directed graph on a finite vertex set \mathcal{S} of size d , with edge set denoted by \mathcal{E}_G . Let M_G denote the incidence matrix of the graph, namely the $d \times d$ nonnegative matrix $M_G = [m(x, y)]$, with $m(x, y) = 1$ if $(x, y) \in \mathcal{E}_G$, and $m(x, y) = 0$ otherwise. Assume that each vertex has at least one out-going edge. For $n \geq 1$ and $x \in \mathcal{S}$, let $N_n(x)$ denote the number of directed paths of length n starting at x . Then the growth rate of the number of directed paths in the graph, namely

$$\max_{x \in \mathcal{S}} \lim_{n \rightarrow \infty} \frac{1}{n} \log N_n(x)$$

exists and equals $\log \rho(M_G)$, where $\rho(M_G)$ is the Perron-Frobenius eigenvalue of M_G .

It is also known that this common limit can be written as

$$\sup_{G\text{-compatible } (\Pi, \pi)} - \sum_{x, y} \pi(x) \pi(y|x) \log \pi(y|x) . \quad (20)$$

Here Π ranges over $d \times d$ transition probability matrices that are G -compatible for the directed graph G , i.e. such that $\pi(y|x) > 0$ implies that $(x, y) \in \mathcal{E}_G$, and π ranges over invariant probability distributions for Π . Note that this is the largest entropy rate among all stationary Markov chains whose transition probability matrix is compatible with the graph.

This characterization of the growth rate of the number of paths in an irreducible graph is a consequence of the Donsker-Varadhan formula for the Perron-Frobenius eigenvalue of a nonnegative matrix. Let us verify this as a corollary of Theorem 4 in the case without controls. We take the state space in Theorem 4 to be \mathcal{S} , i.e. the vertex set of the graph. The control space U is a set consisting of a single point, which we write as $U = \{u\}$. Let $p(y|x, u) := \frac{1}{d(x)}$ for $d(x) :=$ the out-degree of x and $(x, y) \in \mathcal{E}_G$, and let

$$r(x, u, y) := \begin{cases} \log d(x) & \text{if } (x, y) \in \mathcal{E}_G \\ -\infty & \text{otherwise.} \end{cases} \quad (21)$$

Substituting these into the right hand side of (10) gives the expression in (20).

We now bring risk-sensitive control into this mix of ideas. Let U be a finite set and suppose now that for each $u \in U$ we are given a directed graph G_u with vertex set \mathcal{S} . Assume that each vertex has at least one out-going edge in each G_u . We pose the problem of maximizing

$$\max_{x \in \mathcal{S}} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \hat{N}_n(x) ,$$

where now $\hat{N}_n(x)$ is the largest number of directed paths of length n one can create when starting at x and at each time choosing one of the graphs along which to move (i.e. one of the control actions) depending on the history of the states visited so far. More generally, we might allow for a randomized choice of the graph to be used at each time, based on the history of the states and the realizations of the control so far, and ask for the maximum growth rate of the expectation of the number of directed paths of each length that we can create in this way.

This problem can be posed in a framework that is amenable to an application of Theorem 4. As in the case without controls, we set $p(y|x, u) := \frac{1}{d_u(x)}$ for all $(x, y) \in \mathcal{E}_{G_u}$, where $d_u(x)$ denotes the out-degree of vertex x in G_u , and we now set

$$r(x, u, y) := \begin{cases} \log d_u(x) & \text{if } (x, y) \in \mathcal{E}_{G_u} \\ -\infty & \text{otherwise.} \end{cases} \quad (22)$$

According to Theorem 4 this maximum growth rate is given by

$$\max_{\eta} - \sum_{x, u} \tilde{\eta}(x, u) \sum_{y : (x, y) \in \mathcal{E}_{G_u}} \eta_2(y|x, u) \log \eta_2(y|x, u) ,$$

where the maximum is over all $\eta(x, u, y) = \tilde{\eta}(x, u) \eta_2(y|x, u)$ with $\eta_2(y|x, u) > 0$ implying that $(x, y) \in \mathcal{E}_{G_u}$, and such that

$$\sum_{(x, u)} \tilde{\eta}(x, u) \eta_2(y|x, u) = \eta_0(x) ,$$

where, as usual, $\eta_0(x) := \sum_u \tilde{\eta}(x, u)$. Note that this has following interpretation: among all stationary Markov chains $((X_n, Z_n), n \geq 0)$ with state space $\mathcal{S} \times U$ that are compatible with the family of graphs in the sense that if a transition from (x, u) to (y, u') has positive probability then $(x, y) \in \mathcal{E}_{G_u}$,

maximize the conditional entropy of the next state given the current state-entropy pair, i.e. maximize $H(X_1|X_0, U_0)$.

The interpretation of the growth rate of the number of directed paths of a given length in a directed graph as an entropy rate has considerable practical importance in coding theory. Each directed path of length n can be viewed as an *allowed sequence* of length n , with coordinates from the state space \mathcal{S} , and the set of such directed paths is then viewed as a set of *constrained sequences* [14, Problem 4.16], [31]. The problem of *constrained coding* has been extensively studied. In one version of this problem, the goal is to come up with algorithms that can take an infinitely long sequence of symbols from a finite set of size m and produce \mathcal{S} -valued sequences as output in a one-to-one fashion, and such that the output sequences meet the constraints defined by the graph, see [31, Sec. 5.2] for more details. Naturally, it is not possible to do this if $\log m$ exceeds the growth rate given by (20); finding efficient algorithms to do this whenever $\log m$ is less than the growth rate given in (20) was a key early success in this area [1], [31]. Investigating the question of constrained coding up to the maximum possible conditional entropy rate given by the application of Theorem 4 to the controlled graph formulation above would be an interesting challenge.

5.2 Portfolio optimization

As another example, we consider the portfolio optimization problem from [6], except that we consider the reward maximization framework instead of cost minimization as in the classic work of Cover [13]. The model is as follows. The underlying ‘factor process’ $\{X_n\}$ is a discrete time Markov chain on a finite state space $\mathcal{Q} := \{1, \dots, m\}$ (say) with an irreducible transition matrix $Q = [[q(j|i)]]$. The control space will be the simplex $A := \{a = a_1, \dots, a_m \in \mathcal{R}^m : a_i \geq 0 \forall i, \sum_i a_i \leq 1\}$, with a_i denoting the proportion of wealth invested in the i th risky asset. In particular, $1 - \sum_i a_i$ is then the proportion invested in the risk-less bank account. We denote by $\{\pi_n\}$ the A -valued control sequence, representing the trading strategy, i.e., $\pi_{n,i}$ will be the proportion of wealth invested in the i th risky asset at time n . $\{W_n\}$ is the process of m -dimensional vectors of price relatives such that W_{n+1} is conditionally independent of $X_i, i < n; W_i, \pi_i, i \leq n$, given (X_n, X_{n+1}) and its conditional law given the latter is specified by a kernel $\nu(x, y, dw) : \mathcal{Q} \times \mathcal{Q} \mapsto \mathcal{R}^m$ with support in the interior of the positive cone in \mathcal{R}^m . Let $e^r, r > 0$, denote the per period multiplier of wealth invested in the bank account (thus

$e^r - 1$ is the interest rate). Let $\mathbf{1}$ denote the constant vector of all 1's. The evolution of the wealth process $\{V_t\}$ is given by

$$V_{n+1} = V_n [e^r + \langle \pi_n, W_{n+1} - e^r \mathbf{1} \rangle] ,$$

where $V_0 := 1$. The objective is to maximize the risk-adjusted growth rate of wealth

$$\liminf_{n \uparrow \infty} \frac{1}{n} \log E \left[e^{-\frac{\theta}{2} \log V_n} \right] . \quad (23)$$

Here θ is the risk sensitivity parameter. The control sequence $\{\pi_n\}$ is assumed to be adapted to the factor process $\{X_n\}$ and the controls, i.e. the distribution of π_n is chosen as a function of $(X_0, \dots, X_n, \pi_0, \dots, \pi_{n-1})$.

It is useful to contrast the objective we consider with that considered in [6] of maximizing, for $\theta > 0$, the quantity

$$\liminf_{n \uparrow \infty} -\frac{2}{\theta} \frac{1}{n} \log E \left[e^{-\frac{\theta}{2} \log V_n} \right] . \quad (24)$$

In [6], this problem is considered by writing the objective in (24) as

$$-\limsup_{n \uparrow \infty} \frac{2}{\theta} \frac{1}{n} \log E \left[e^{-\frac{\theta}{2} \log V_n} \right] ,$$

and then studying the risk-sensitive cost minimization problem corresponding to the objective

$$\limsup_{n \uparrow \infty} \frac{2}{\theta} \frac{1}{n} \log E \left[e^{-\frac{\theta}{2} \log V_n} \right] .$$

That positive θ indicates risk aversion in (24) is argued, see [7, Eqn. (2.1)], by writing the Taylor's series expansion, for small θ ,

$$-\frac{2}{\theta} \log E \left[e^{-\frac{\theta}{2} \log V_n} \right] = E[\log V_n] - \frac{\theta}{4} \text{var}(\log V_n) + O(\theta^2) .$$

By contrast, our formulation is able to handle both the case of risk-aversion and risk-seeking. The Taylor's series expansion

$$\log E \left[e^{-\frac{\theta}{2} \log V_n} \right] = -\frac{\theta}{2} E[\log V_n] + \frac{\theta^2}{8} \text{var}(\log V_n) + o(\theta^2)$$

indicates that if the objective in (23) is multiplied by $-\frac{2}{\theta}$, then it corresponds to risk-aversion for positive θ and to risk-seeking for negative θ .

Keeping in mind that $e^r + \langle a, W - e^r \mathbf{1} \rangle > 0$ under our assumption on the support of $\nu(x, y, dz)$, define

$$\begin{aligned}\mu(x, a, y) &:= \int e^{-\frac{\theta}{2} \log[e^r + \langle a, w - e^r \rangle]} \nu(x, y, dw), \\ &\quad (\text{assumed to be } < \infty) \\ r(x, a) &:= -\frac{2}{\theta} \log \left(\sum_y q(y|x) \mu(x, a, y) \right), \\ p(y|x, a) &:= \frac{q(y|x) \mu(x, a, y)}{\sum_{y'} q(y'|x) \mu(x, a, y')}.\end{aligned}$$

One can show that for all $n \geq 1$ and all admissible controls, we have

$$\frac{1}{n} \log E \left[e^{-\frac{\theta}{2} \log V_n} \right] = \frac{1}{n} \log \tilde{E} \left[e^{-\frac{\theta}{2} \sum_{m=0}^{n-1} r(X_m, \pi_m)} \right],$$

where \tilde{E} is the expectation with respect to the law

$$\begin{aligned}&p(x_0) \phi_0(da_0|x_0) p(x_1|x_0, a_0) \phi_1(da_1|x_0, a_0, x_1) \dots \\ &\times \phi_{n-1}(da_{n-1}|(x_i, a_i, 0 \leq i \leq n-2), x_{n-1}),\end{aligned}$$

where $p(x_0)$ is the initial distribution of X_0 , the admissible controls are determined by the kernels $\phi_0(\cdot|\cdot), \dots, \phi_{n-1}(\cdot|\cdot)$, and the salient point is that the transition kernel for the evolution of the factor process under this change of measure is given by the kernel $p(\cdot|\cdot, \cdot)$ defined above. To see this, first observe that W_1, \dots, W_n are conditionally independent and identically distributed given $(X_i, \pi_i, 0 \leq i \leq n)$. Hence

$$\begin{aligned}E \left[e^{-\frac{\theta}{2} \log V_n} | X_i, \pi_i, 0 \leq i \leq n \right] &= E \left[\prod_{m=0}^{n-1} e^{-\frac{\theta}{2} \log \frac{V_{m+1}}{V_m}} | X_i, \pi_i, 0 \leq i \leq n \right] \\ &= \prod_{m=0}^{n-1} E \left[e^{-\frac{\theta}{2} \log \frac{V_{m+1}}{V_m}} | X_i, \pi_i, 0 \leq i \leq n \right] \\ &= \prod_{m=0}^{n-1} \mu(X_m, \pi_m, X_{m+1}),\end{aligned}$$

so we have

$$E \left[e^{-\frac{\theta}{2} \log V_n} \right] = E \left[\prod_{m=0}^{n-1} \mu(X_m, \pi_m, X_{m+1}) \right] .$$

For an admissible control strategy, we can write this as

$$\sum_{x_0, \dots, x_n} \int_{a_0} \dots \int_{a_{n-1}} p(x_0) \prod_{m=0}^{n-1} \mu(x_m, a_m, x_{m+1}) q(x_{m+1} | x_m) \\ \phi_m(da_m | (x_i, a_i, 0 \leq i \leq m-1), x_m) ,$$

which is the same as

$$\sum_{x_0, \dots, x_n} \int_{a_0} \dots \int_{a_{n-1}} p(x_0) \prod_{m=0}^{n-1} e^{-\frac{\theta}{2} r(x_m, a_m)} p(x_{m+1} | x_m, a_m) \\ \phi_m(da_m | (x_i, a_i, 0 \leq i \leq m-1), x_m) ,$$

which equals $\tilde{E} \left[e^{-\frac{\theta}{2} \sum_{m=0}^{n-1} r(X_m, \pi_m)} \right]$.

Hence the problem of maximizing (23) is equivalent to the risk-sensitive control problem for a controlled Markov chain on \mathcal{Q} with action space A and controlled transition probabilities $p(y|x, a)$, $x, y \in \mathcal{Q}$, $a \in A$, the objective being to maximize the reward

$$\lambda := \sup_{x_0} \sup \liminf_{n \uparrow \infty} \frac{1}{n} \log E \left[e^{-\frac{\theta}{2} \sum_{m=0}^{n-1} r(X_m, \pi_m)} | X_0 = x_0 \right] .$$

where the second supremum is over admissible controls.

The optimal growth rate for the wealth is then given by

$$\lambda = \max_{\eta \in \mathcal{G}} \left(\sum_x \int_A \tilde{\eta}(x, da) \left(-\frac{\theta}{2} r(x, a) - \sum_y \int_A \eta_2(y|x, a) \log \left(\frac{\eta_2(y|x, a)}{p(y|x, a)} \right) \right) \right)$$

where

$$\mathcal{G} := \left\{ \eta(x, da, y) \in \mathcal{P}(\mathcal{Q} \times A \times \mathcal{Q}) : \eta(x, da, y) = \tilde{\eta}(x, da) \eta_2(y|x, a) \right. \\ \left. = \eta_0(x) \eta_1(da|x) \eta_2(y|x, a) \text{ such that } \eta_0 \text{ is stationary under} \right. \\ \left. \text{the transition matrix } \left[\left[\int \eta_1(da|x) \eta_2(y|x, a) \right] \right] \right\} .$$

In order to justify this, we need to verify that the conditions **(A0)** and **(A1)** are satisfied. Here \mathcal{Q} plays the role of \mathcal{S} , A plays the role of U , and $-\frac{\theta}{2}r(x, a)$ plays the role of $r(x, u, y)$ in the general theory. The validity of **(A0)** follows from the continuity of the logarithm function. The validity of **(A1)** follows from the continuity of the logarithm function, the fact that \mathcal{Q} is finite, and because $\sum_{y'} q(y'|x)\mu(x, a, y)$ is strictly positive for all (x, a) .

If we discretize A , this is a finite dimensional concave maximization problem eminently amenable to standard nonlinear programming tools.

5.3 Minimizing exit rate from a domain

Consider a set of controlled stochastic matrices on a finite state space $S = \{1, \dots, s\}$ denoted by $P_u = [[p(j|i, u)]]_{i,j \in S}$. Here u is the control parameter taking values in A , where A is a compact metric action space. We assume that $u \mapsto P_u$ is continuous and P_u is irreducible for all u . Let $S_0 \subset S$ be a nonempty proper subset of S and let $S_1 := S_0^c$ denote its complement. Let \check{P}_u denote the restriction of P_u to S_1 and for a sequence of random variables $\{X_n\}$ with values in S , define $\tau := \inf\{n \geq 0 : X_n \in S_0\}$.

We are interested in determining

$$\lambda := \sup_{i \in S_1} \sup \liminf_{n \uparrow \infty} \frac{1}{n} \log P(\tau > n) ,$$

where the second supremum is over all admissible controls, and the law of τ is determined by the control strategy. Namely, we are interested in the problem of finding the slowest exit rate from S_1 over admissible control strategies.

Write $\check{P}_u = D_u Q_u$ where D_u is a diagonal matrix with its i th diagonal entry $d(i, u) := \sum_{j \in S_1} p(j|i, u)$ and $Q_u := [[q(j|i, u)]]$ is a stochastic matrix on S_1 given by $q(j|i, u) := d(i, u)^{-1} p(j|i, u)$, where we will also assume that $d(i, u) > 0$ for all $i \in S_1$ and $u \in A$. It can be checked that for any admissible control strategy and $i \in S_1$, we have

$$P(\tau > n) = E \left[e^{\sum_{m=0}^{n-1} \log(d(X_m, U_m))} \right] ,$$

where U_m denotes the choice of control at time m , and $\{X_n\}$ is the S_1 -valued Markov chain, having the transition probability matrix Q_{U_m} at time m . Therefore, with the choices $\mathcal{S} := S_1$, $U := A$, and $r(i, u, j) := \log d(i, u)$, the problem is amenable to our general theory.

Disintegrate a typical element $\eta \in \mathcal{P}(S_1 \times A \times S_1)$ as $\eta_0(i)\eta_1(du|i)\eta_2(j|i, u)$, and write $\tilde{\eta}(i, du)$ for $\eta_0(i)\eta_1(du|i)$.

Then our results show that

$$\lambda = \max_{\eta \in \mathcal{G}} \left(\sum_{i,j \in S_1} \int_A \eta(i, du, j) \log(d(i, u)) - \sum_{i \in S_1} \int_A \tilde{\eta}(i, du) D(\eta_2(j|i, u) || q(j|i, u)) \right),$$

where \mathcal{G} denotes the set of $\eta \in \mathcal{P}(S_1 \times A \times S_1)$ for which η_0 is invariant under the transition kernel $\int_A \eta_1(du|i)\eta_2(j|i, u)$. To verify this, we need to check the validity of the conditions **(A0)** and **(A1)**. The former is a consequence of the assumed continuity of $u \mapsto P_u$. The latter is a consequence of the fact that S_1 is finite and that $u \mapsto Q_u$ is continuous, which in turn follows from the assumed continuity of $u \mapsto P_u$ and the assumption that $d(i, u) > 0$ for all $i \in S_1$ and $u \in A$.

6 Concluding remarks

We considered the problem of maximizing the growth rate of reward in the standard risk-sensitive formulation for a controlled Markov chain on a compact metric state space, with a compact metric action space. We took a non-standard approach to this problem via a nonlinear version of the Krein-Rutman theorem to obtain a variational formulation for the optimal reward. This leads to an occupation measure based concave maximization formulation of the control problem.

The approach holds promise for possible use of convex optimization techniques for approximate solution of the risk-sensitive reward maximization problem, in a manner analogous to what abstract linear programming does for the classical additive reward problems (such as discounted or ergodic rewards, see, e.g., [25]). We achieved this with rather few technical conditions except for the compactness of the state and action spaces. It remains a major challenge to extend this approach to noncompact state and action spaces.

References

- [1] Adler, R. L., Coppersmith, D., and Hassner, M. (1983) “Algorithms for sliding block code: an application of symbolic dynamics to information theory”, *IEEE Trans. Information Theory* 29(1), 5-22.
- [2] Aliprantis, C. D., and Tourky, R. (2007) *Cones and Duality*, Graduate Studies in Mathematics, Vol. 84, American Mathematical Society, Providence, RI, USA.
- [3] Arapostathis, A. (2013) “A correction to Mahadevan’s nonlinear Krein-Rutman theorem”, (preprint).
- [4] Arapostathis, A., Borkar, V. S., and Kumar, K. S. (2013) “Risk-sensitive control and an abstract Collatz-Wielandt formula”, arXiv preprint arXiv:1312.5834.
- [5] Beneš, V. E. (1970) “Existence of optimal strategies based on specified information, for a class of stochastic decision problems”, *SIAM J. Control* 8(2), 179-188.
- [6] Bielecki, T., Hernandez-Hernández, D., and Pliska, S. R. (1999) “Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management”, *Math. Methods of Op. Research* 50(2), 167-188.
- [7] Bielecki, T., Pliska, S. R. and Sherris, M. (2000) “Risk sensitive asset allocation”, *J. Econ. Dyn. and Control* Vol. 24, 1145 -1177.
- [8] Bierkens, J., and Kappen, H. J. (2014) “Explicit solution of relative entropy weighted control”, *Systems & Control Letters* 72, 36-43.
- [9] Billingsley, P. (1968) *Convergence of Probability Measures*, John Wiley & Sons, New York, USA.
- [10] Borkar, V. S. (1989) *Optimal Control of Diffusion Processes*, Pitman Research Notes in Math. No. 203, Longman Scientific & Technical, Harlow, UK.
- [11] Borkar, V. S. (2002) “Convex analytic methods in Markov decision processes”, in ‘*Handbook of Markov Decision Processes*’ (E. A. Feinberg and A. Shwartz, eds.), Kluwer Academic Publishers, Boston, 347-375.

- [12] Chang, K. C. (2009) “A nonlinear Krein Rutman theorem”, *J. Systems Sci. and Complexity* 22(4), 542-554.
- [13] Cover, T. M. (1991) “Universal portfolios”, *Math. Finance* 1(1), 1-29.
- [14] Cover, T. M., and Thomas, J. A. (2006) *Elements of Information Theory* (2nd ed.), Wiley-Interscience, New Jersey, USA.
- [15] Dembo, A., and Zeitouni, O. (1998) *Large Deviations Techniques and Applications* (2nd ed.), Springer Verlag, New York, USA.
- [16] Di Masi, G. B., and Stettner, L. (2007) “Infinite horizon risk sensitive control of discrete time Markov processes under minorization property”, *SIAM J. Control and Optim.* 46(1), 231-252.
- [17] Dupuis, P. and Ellis, R. S. (1997) *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley, New York.
- [18] Donsker, M. D., and Varadhan, S. R. S. (1975) “On a variational formula for the principal eigenvalue for operators with maximum principle”, *Proc. Nat. Acad. Sci. USA* 72(3), 780-783.
- [19] Fan, K. (1952) “Fixed point and minimax theorems in locally convex topological linear spaces”, *Proc. Nat. Acad. Sci. USA* 38, 121-126.
- [20] Fleming, W. H., and Hernandez-Hernández, D. (1996) “Risk-sensitive control of finite state machines on an infinite horizon I”, *SIAM J. Control and Optim.* 35(5), 1790-1810.
- [21] Fleming, W. H., and Hernandez-Hernández, D. (1996) “Risk-sensitive control of finite state machines on an infinite horizon II”, *SIAM J. Control and Optim.* 37(4), 1048-1069.
- [22] Friedland, S. (1990) “Characterizations of spectral radius of positive operators”, *Linear Algebra and its Applications* 134, 93-105.
- [23] Guan, P., Raginsky, M., and Willett, R. M. (2014) “Online Markov decision processes with Kullback-Leibler control cost”, *IEEE Trans. Automatic Control* 59(6), 1423-1438.

- [24] Hernandez-Hernández, D., and Marcus, S. I. (1996) Risk sensitive control of Markov processes in countable state space”, *Systems and Control Letters* 29(3), 147-155.
- [25] Hernández-Lerma, O., and Lasserre, J. B. (2002) “The linear programming approach”, in ‘*Handbook of Markov Decision Processes*’ (E. A. Feinberg and A. Shwartz, eds.), Kluwer Academic Publishers, Boston, 377-407.
- [26] Jaśkiewicz, A. (2007) “Average optimality for risk-sensitive control with general state space”, *Ann. Appl. Prob.* 17(2), 654-675.
- [27] Jaśkiewicz, A. (2007) “A note on risk-sensitive control of invariant models”, *Systems & Control Letters* 56(11-12), 663-668.
- [28] Jaśkiewicz, A. (2008) “A note on negative dynamic programming for risk-sensitive control”, *Operations Research Letters* 36(5), 531-534.
- [29] Krein, M. G., and Rutman, M. A. (1948) “Linear operators leaving invariant a cone in a Banach space”, *Uspekhi Mat. Nauk*, 3(1:23), 3-95.
- [30] Lemmens, B., and Nussbaum, R. (2012) *Nonlinear Perron-Frobenius Theory*, Cambridge Tracts in Mathematics, Vol. 189, Cambridge University Press, Cambridge, UK.
- [31] Lind, D., and Marcus, B. (1995) *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, UK.
- [32] Mahadevan, R. (2007) “A note on a non-linear Krein-Rutman theorem”, *Nonlinear Analysis: Theory, Methods & Appl.* 67(11), 3084-3090.
- [33] MathSciNet review of [37] by R. D. Nussbaum, MR1338356.
- [34] Meyer, C. (2000) *Matrix Analysis and Applied Linear Algebra*, SIAM.
- [35] Nisio, M. (1978) “On stochastic optimal controls and envelope of Markovian semigroups”, in *Proc. Intl. Symp. on Stochastic Differential Equations, RIMS, Kyoto Uni., Kyoto, 1976* (K. Ito, ed.), Wiley, New York, 297-325.

- [36] Nussbaum, R. D. (1980) “Eigenvalues of nonlinear positive operators and the linear Krein-Rutman theorem”, in *Lec. Notes in Math.*, Vol. 886, Springer, Berlin, 1981, 309-330.
- [37] Ogiwara, T. (1995) “Nonlinear Perron-Frobenius problem on an ordered Banach space”, *Japan J. Math.*, 21, 43 -103.
- [38] Pollard, D. (2002) *A User’s Guide to Measure Theoretic Probability*, Cambridge Uni. Press, Cambridge, UK.
- [39] Rudin, W. (1973) *Functional Analysis*, McGraw-Hill, New York, USA.
- [40] Todorov, E. (2007) “Linearly-solvable Markov decision problems”, *Proc. Advances in Neural Information Processing Systems, Vol. 19* (B. Schölkopf, J. Platt and T. Hoffman, eds.), MIT Press, Cambridge, Mass., 2007, 1369-1376.
- [41] Van Erven, T., and Harremöes, P. (2014) “Rényi divergence and the Kullback-Leibler divergence”, *IEEE Trans. Information Theory*, 60(7), 3797 -3820.